

DOCUMENT RESUME

ED 322 773

FL 018 789

AUTHOR Kerswill, Paul; Wright, Susan
TITLE On the Limits of Auditory Transcription: A Sociophonetic Approach.
PUB DATE 89
NOTE 27p.; In: York Papers in Linguistics 14; see FL 018 786.
PUB TYPE Reports - Research/Technical (143)
EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Auditory Perception; Dialect Studies; Language Research; Linguistic Theory; *Phonetics; Sociolinguistics
IDENTIFIERS *Transcription

ABSTRACT

A study examined what trained phoneticians do when they are presented with a transcription task to carry out without any knowledge of the dialect they are listening to and without any explicit phonological theory as a point of departure. The "best" tokens of three categories of potential assimilation (full, partial, and zero alveolar) were drawn from electropalatograms and recordings of sentences and recordings made by a linguistically naive speaker. These tokens were then transferred to a test tape and transcribed by 13 phoneticians, who provided a narrow transcription, lexical identification, and rating by category. Results suggest that identification of these tokens as alveolar or otherwise involves an extremely complex set of factors, including knowledge of the auditory effects of different tongue gestures, differential vowel quality and length in different consonantal contexts, and dialect-specific allophonic differences. The task of transcription is seen as an inherently ambiguous task at several levels, and transcription without any kind of theory is not recommended. (MSE)

* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *

ON THE LIMITS OF AUDITORY TRANSCRIPTION:
A SOCIOPHONETIC APPROACH

Paul Kerswill

"PERMISSION TO REPRODUCE THIS
MATERIAL HAS BEEN GRANTED BY

S. Harlow

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

- ☒ This document has been reproduced as
received from the person or organization
originating it.
- ☐ Minor changes have been made to improve
reproduction quality.
- Points of view or opinions stated in this docu-
ment do not necessarily represent official
OERI position or policy.

ED322773

FL018789

ON THE LIMITS OF AUDITORY TRANSCRIPTION: A SOCIOPHONETIC APPROACH*

Paul Kerswill

Department of Linguistic Science, University of Reading

Susan Wright

Department of Linguistics, University of Cambridge

1. Introductory remarks

In a way, this paper only obliquely addresses sociolinguistic issues. But we can justify its inclusion in this volume in that it fits into a way of thinking that has been characteristic of a number of sociolinguists in the last eight or ten years. During this period, sociolinguistics has become something of a self-scrutinising subject, in that people have questioned not only the methodology but also the linguistic and social theory behind it. This paper can be seen as a contribution to this discussion.

However, it intends to do so in a novel way. We will tackle an area of sociolinguistic methodology which is rarely discussed; and we are going to try to show that this is of no less theoretical significance: this is the phonetic nature, and linked with that the transcription, of the actual sounds uttered by speakers. At first sight, this seems to be a purely atheoretical problem, a matter of nuts and bolts. After all, transcription is something that, with a bit of ear-training, we can all get reasonably good at; as such, it's simply a tool of the trade. But in our view that is

* Authors' correspondence addresses: Department of Linguistic Science, University of Reading, PO Box 218, Whiteknights, Reading RG6 2AA (Kerswill) and Department of Linguistics, University of Cambridge, Faculty of Modern and Medieval Languages, Sidgwick Avenue, Cambridge, CB3 9DA (Wright).

not all that transcription is about; it is also part of the theory of sociolinguistics.

There are two reasons why we think this is so. Firstly, from the speakers' point of view, the sounds are what they use to convey complex indexical information. Secondly, from the point of view of the linguists, for them to do their transcription they need a phonological theory, however rudimentary. Without a theory, they cannot know what kind of detail to transcribe, and with the wrong theory they will transcribe the wrong detail. Towards the end of this paper, we will show what trained phoneticians do when they are presented with a transcription task to carry out 'cold', without any knowledge of the dialect they are listening to, and without any explicit phonological theory as a point of departure.

In fact, quite a lot of attention has been paid to the LINGUISTIC representation of the variants of phonological variables, notably by Knowles (1978), Lodge (1986), the Milroys (e.g. J. Milroy, 1976) and Harris (e.g. 1986). Regrettably (for reasons that will become clear), this has not gone hand-in-hand with a consideration of what happens during the act of transcribing those variants. This will be the central concern of this paper.

2. The importance of phonetic transcription

Before we look at the experiments we carried out, we will consider in more detail WHY it is important to examine phonetic transcription. We will approach this question from two angles: first, from the point of view of recent dialectology and sociolinguistics in general; and second, from the specific point of view of a more phonetics-based field which can be termed 'sociophonetics' - in particular, our own work on connected speech processes in local Cambridge English.

2.1 Dialectology and sociolinguistics

First, then, some general points about dialectology and sociolinguistics.

We can start with a rather alarming discovery made a few years ago by Peter Trudgill (1983: 31-51). On the basis of the *Linguistic Atlas of England* (Orton et al., 1978), he drew the map shown in Fig. 1.

FIGURE 1
Vowel in 'last' (from Trudgill, 1983)



This shows the reflexes of Middle English *a* in the word *last*. The point to note is this. There is a large band across the middle of England where the vowel is [a:] which separates two areas with [ɑ:], one to the south-east, the other to the north-east, in Norfolk. Trudgill was suspicious of this transcription, believing a front vowel to be more usual in Norfolk.

He noted that this Norfolk [ɑ:] area (marked with an arrow) in fact covers the locations surveyed by one particular field-worker, who quite simply 'got it wrong' (1983: 40); the result is a 'field-worker isogloss' (op. cit.: 38). The moral here is obvious.

Sociolinguists and dialectologists have relied heavily on auditory phonetic transcription as a basic analytical tool in their investigation of variation and sound change. And, as we mentioned, it has been treated by them as a pre-theoretical notion, and they have regarded it as a tedious but necessary evil. In most of the early studies, little attention was paid to the transcription itself, though the precise effect of this failure (if this is the appropriate word for it) is hard to assess. There are two important issues here. These are, first, the *reliability* and, second, the *validity* of the transcriptions. First, let us look at reliability: how consistent are transcriptions both across transcribers and within transcribers? The more significant of these, we think, is within-transcriber variability, since most of the transcription is usually done by a single person. The main question here is whether or not a transcriber is consistent: will he or she transcribe the same token the same way twice? And does that transcriber have a tendency to 'drift' in his or her judgments over a period of time? We shall not in this paper have any more to say on the subject of reliability. We shall be more concerned with the validity of the transcriptions. Here, the main question is the way in which a transcription reflects (a) articulatory facts and (b) auditory impressions. So we might like to consider whether there is a consistent bias towards a particular transcription in, say, a particular phonological environment, or whether manner of articulation influences the perception of place of articulation.

We mentioned earlier the increasing sophistication of the linguistic variable. Just to give some idea of how complex a variable can become, consider Table 1, which shows one linguist's analysis of the variants of two vowels in Liverpool:

TABLE 1

/uə/ and /ɔə/ in Liverpool (from Knowles, 1978: 85)

		<i>sure</i> /uə/	<i>shore</i> /ɔə/
1	Lax [u, ɔ] before an unstressed vowel: or	ʊə	ɔə
2	(a) diphthongize [u, ɔ]: or	ʊé, ʊé	ɔuə
	(b) front [u]:	ʌə	
3	modify VVV to V + glide + V:	ɪwə	ɔwə
4	front final [ə]:	uɛ, ?ʊɛ ʊé, ʊuɛ ʌɛ, ɪwɛ	oɛ, ?ɔɛ ɔuɛ ɔwɛ

He sees these variants as generated by a set of interacting rules, which represent 'the options open to the speaker at different stages in speech production, and the way these options can be used to convey sociolinguistic information about the speaker' (1978: 90). Similarly, for the consonantal variable (ng), corresponding to the velar nasal in RP *sing*, Knowles identifies the following variants, again generated by rules (op. cit.: 86):

sɪŋ	sɪŋ
sɪŋːg	sɪːŋ
sɪŋː	sɪːŋ ^g

Knowles' analysis is multidimensional. This is true also of the Milroys' analysis of Belfast vowels. Table 2 (taken from Milroy, 1987: 124) shows the variables (o) and (ɛ), as realised in the data for a single speaker. Milroy argues that these variants should be analysed in terms of three sub-scales: roundness, backness and length.

Clearly, if sociolinguists are going to operate with this amount of detail, we need to know something about the reliability and validity of the transcriptions on which their (usually sophisticated) analyses are based.

TABLE 2
(o) and (ε) in Belfast (from Milroy, 1987: 124)

(o):	ɑ	ɑ	ɑː	ɔ	ɔː
	got		shop		
	Polytech				
	shop			probably	job
	pot	concentrated			of
		vodka		bottom	God
(ε):	ɛ	ɛː	ɛ	ɛː	
	set-up		specials	red	
	lent		went	tell	
	went			ten	
	specials				
	remember				
	twenty				

2.2 Sociophonetics

If Liverpool causes difficulties for the transcriber, this is even more true of a relatively new field of study, which intersects, to a greater or lesser extent, with correlational sociolinguistics. This is the growing field of sociophonetic research. A recent, though largely descriptive example is Lodge's (1986) outline of the phonetics and phonologies of a number of non-standard varieties of English. In it, he pays special attention to the word in connected speech. As in other sociolinguistic studies, Lodge uses an auditory transcription, noting quite fine detail.

Lodge is not only interested in 'traditional' phonetic variables, but also in the range of assimilations, deletions and epentheses of normal connected speech. In our work in Cambridge, we too have focused on these phenomena, which we call *connected speech processes*, or *CSPs*.

We will digress a little at this point to say something about the background to our Cambridge project, so as to make it plain just why we have conducted the transcription experiments we are going to be reporting. Unlike the 'traditional' variables of sociolinguistics, CSPs are in some sense phonetically motivated: that is, their application can be explained with reference to the physiology and the dynamics of the vocal tract. Our own interest in these phenomena derives from two sources. The first is the observation that conditioned sound changes are always the result of the fossilisation of CSPs. The second concerns the fact that CSPs tend, despite their 'naturalness', to be to some extent variety-specific. This is shown in Dressler's work in Vienna (Dressler & Wodak, 1982) and in the work of one of us in Durham (Kerswill, 1987). Dressler talks about lenition and fortition rules (i.e. CSPs) which serve to ease production (in the case of lenitions) or to ease perception (in the case of fortitions). Some of these processes are apparently specific to one or other of the two major varieties of German spoken in Vienna: the local dialect and standard Austrian German. In Durham, Kerswill observed that certain processes usually described for English appeared to be absent, while others not generally found in the literature were present. The two clearest examples are those shown in Table 3, overleaf. By combining the facts of sound change and the variety-specific nature of CSPs with the sociolinguistic axiom that sounds undergoing change are sociolinguistically salient, we arrived at the basic hypothesis of our study. This is, to put it quite simply, that some connected speech processes will show social differentiation in a speech community.

TABLE 3

Connected speech processes in Durham (from Kerswill, 1987: 44)

(1) CSP present in Durham, absent in RP:

Regressive voicing assimilation:

like [g] baims;
 like [g] me;
 each [dʒ] deputy;
 this [z] village;
 scraped [d] down;
 what's [dʒ] gone in, man?
 good chap [b], Jack

(2) CSP present in RP, absent in Durham:

Assimilation of place of articulation:

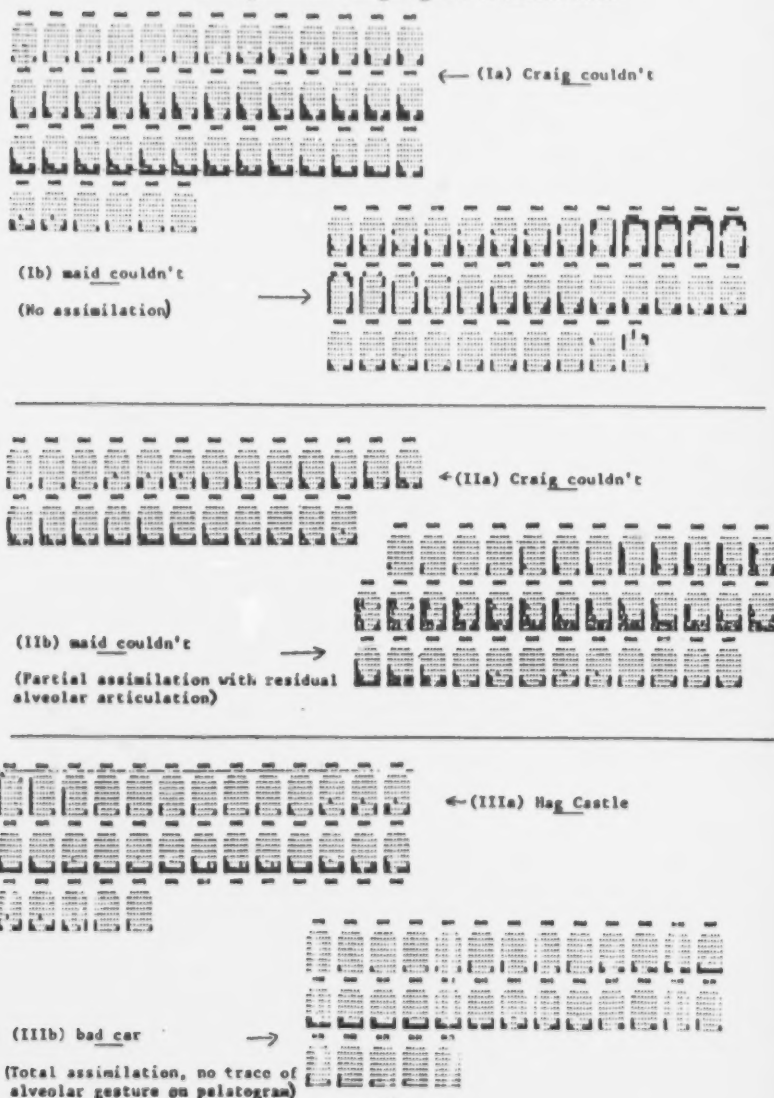
that pen [ðæʔt pɛn] -> [ðæʔppɛn]
 that cup [ðæʔt kʌʔp] -> [ðæʔkkʌʔp]
 good pen [gɒd pɛn] -> [gɒbpɛn]
 good car [gɒd kɑ:] -> [gɒgkɑ:]

In the Cambridge project, we are looking at a range of processes, particularly place assimilation, /-vocalisation, syllable deletion and palatalisation. We are doing this combining the techniques of sociolinguistics with those of experimental phonetics. We are looking at natural speech from a sample of speakers differentiated by social class, sex and age. At the same time as looking at social differentiation, we are also looking at the effects of speech style, particularly speaking rate, as well as the more usual style parameter of formality. (Various aspects of the project, along with some of our results, are reported in Kerswill, 1985b; Wright, 1986; Nolan and Kerswill, 1988; Wright and Kerswill, 1989.)

An important part of our hypothesis is that some CSPs will behave in a way comparable with 'ordinary' sociolinguistic variables. We also hypothesise that some of these sociolinguistically salient CSPs will tend towards articulatory discreteness: that is, they will apply in an all-or-nothing way. They will, in other words, be beginning to show the characteristics of fossilisation and subsequent phonologisation. On the other hand, non-salient CSPs will be more purely phonetically, or naturally, motivated, and will be directly sensitive to speaking rate changes. As such, we can expect them to be phonetically gradual in their application. We can, then, expect to find varying degrees of partial deletions, partial assimilations, residual articulatory gestures, etc. This notion of articulatory gradualness would seem to be especially relevant to one particular favourite sociolinguistic variable: that of final *t* or *d* deletion; yet gradualness does not appear to have been considered in the context of these variables.

We need, then, to be able to identify this articulatory gradualness. To do this, we carried out an electropalatographic study of assimilations. *Electropalatography* (EPG) is a technique which allows the dynamic contact of the tongue against the roof of the mouth to be recorded. The subject wears a specially-made acrylic palate in which are embedded 62 electrodes. A computer records the contact of the tongue with these electrodes. Fig. 2 shows some typical EPG output. Each 'palatogram' shows the degree of lingual contact with the palate during a particular 10 ms window; the top row of dots represents electrodes situated along the alveolar ridge, the bottom row those at the junction between the hard and soft palates. Fig. 2 (overleaf) shows the tongue contacts at the word boundaries in utterances where there is a (potential) assimilation of a final *d* to an initial *k*, (Ib, IIb, IIIb) together with 'control' utterances with 'underlying' final *g* (Ia, IIa, IIIa). Details of the analysis will be given below; but suffice it to say that there is clear evidence here of articulatorily gradualness, shown by the progression from a complete lack of assimilation (Ib), through a partial assimilation (IIb), to a complete assimilation (IIIb).

FIGURE 2
Palatograms showing degrees of assimilation



From the point of view of transcription, the relationship between articulation and the percept is an extremely important one. This is not only true with gradual processes such as assimilation, but also in transcription generally. To illustrate this, we can take the potential minimal pair shown below. Do these ever merge, as suggested by the transcription given, or will there always be some articulatory or auditory difference?

fang collector, fan collector -> [fæŋ kəlektə]

The question is: does perception operate phoneme-categorially, and classify intermediate forms decisively as (in this example) *fan* or *fang*; and if so, can we talk in terms of a perceptual boundary lying on a putative continuum of alveolar loss? How would this affect a phonetician's attempt at transcribing a potential assimilation? The relationship between articulation and perception is something that our experiment has tried to elucidate.

Finally, before we consider the experiment, we shall raise an issue that is well known, but still not sufficiently discussed: this is the likelihood that a segmental transcription predisposes the phonetician to transcribe a series of discrete articulations, whereas we know that articulations blend and overlap in a complex way. It is true that a transcription can record double articulations, partially overlapping articulations, and even the spread of a feature, such as nasalisation, over more than one segment. Despite this, the segments do get transcribed in sequence. Moreover, and this is important from our point of view, the segments tend to get transcribed in an all-or-nothing way. All this predisposes the transcriber to hear a series of discrete, completely articulated segments.

3. The experiment

This experiment explores the relationship between auditory phonetic transcription and some aspects of articulatory fact by comparing transcriptions of potential assimilations with EPG records of the same tokens.

On the basis of earlier experiments, we decided to use as categories three 'degrees' of assimilation. These are associated with three different scores, and correspond to the categories shown in Fig. 2, above. The categories, or *EPG conditions*, can be more explicitly defined as follows:

1. *Full alveolar*: the EPG record shows a complete alveolar closure at some point during the articulation.
2. *Partial alveolar* (residual alveolar gesture): the record shows more lateral and/or alveolar contact than the non-assimilating environment, but nonetheless shows no complete closure at any point during the articulation.
3. *Zero alveolar* (complete assimilation): the record is either identical with the non-assimilating environment, or else shows less lateral and/or alveolar contact than it.

For reasons which will become clear below, we added a fourth EPG condition:

4. *Non-alveolar* (underlying velar or bilabial).

We then made a list of sentences containing possible word-final assimilations of /d/ to a following velar or bilabial, together with 'control' sentences with underlying velars and bilabials. (We did not include /t/: final /t/ is normally realised as a pure glottal stop in many varieties of English, particularly preconsonantly, as in these examples. Our parallel study of final /n/ will be reported elsewhere, Wright and Kerswill, in prep.) The assimilation 'sites' and their controls are given below:

	Assimilation site	Control
d+k	road collapsed	rogue collapsed
d+k	Byrd concert	Berg concert
d+k	fad catch	fag catch
d+g	did gardens	dig gardens

THE LIMITS OF AUDITORY TRANSCRIPTION

d+g	bed girls	beg girls
d+g	lead got	leg got
d+m	bride must	bribe must

We got a phonetician to make EPG and audio recordings of these sentences. The ones with underlying alveolars were recorded with each of the three 'degrees' of assimilation: with full alveolar articulation, with partial alveolar closure, and with no alveolar closure. The control utterances were also recorded. This gave us tokens of our four 'EPG conditions' - three underlyingly alveolar, one underlyingly velar or bilabial. In all the tokens, any hint of an audible release was avoided. In order to partially guard against any unrepresentativeness in the production by the phonetician, we compared his EPG records with those produced by a linguistically naïve speaker in an earlier experiment. On the basis of this comparison, we picked out the 'best' tokens of each category for use in the listening test.

The tokens were transferred to a test tape in such a way that the 'control' member of each sentence pair occurred four times and each of the three degrees of assimilation for the underlying alveolars occurred twice each. This gave us a tape on which one-third of the tokens were control items. They were ordered such that identical sentences and 'articulation types' were not adjacent. Thirteen other phoneticians then acted as subjects. Their task was to provide the following:

- a narrow transcription (of preceding vowel and consonant assimilation site)
- lexical identification (judgment of underlying final /g,b/ vs. underlying /d/)
- rating of words judged to end in an alveolar as having:

full alveolar contact,
partial alveolar contact, or
zero alveolar contact

In this way, we hoped to be able to see what criteria, if any, the tran-

scribers used in deciding between the various degrees of assimilation and the non-alveolar control environments.

3.2 Articulatory gradualness reflected in identifications

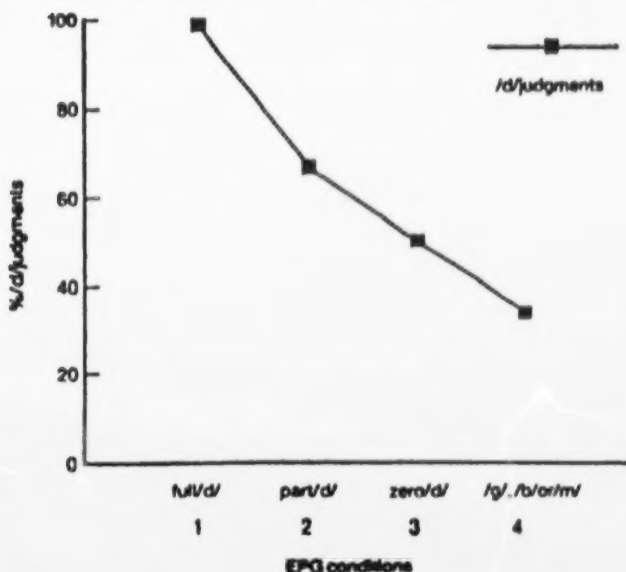
The results of the identification part of the task were as shown in Fig. 3, which gives the phoneticians' judgments of the tokens as underlyingly alveolar. Tokens which were articulated with a complete alveolar closure ('EPG condition 1') were almost consistently identified as alveolar. The percentage of alveolar identifications rapidly drops across the other three EPG conditions - partial alveolar, completely assimilated ('zero') alveolar and underlyingly velar/bilabial.

As we would expect, a good deal of 'alveolarity' seems also to be cued by the auditory impression made by the partially assimilated tokens (condition 2). However, perhaps the most interesting results concern conditions 3 and 4, both of which show substantial alveolar scores. Before attempting to interpret the scores for these two conditions, we should first ask why any of the condition 4 tokens should be rated as alveolar at all. Three factors should be noted: (1) any 'error' will raise, not lower the score; (2) we can expect listeners to try to 'hear' alveolarity even when there is none; and (3) due to redundancy, some phonetic indeterminacy is tolerated in natural speech; here, in the absence of redundancy, the phonetic indeterminacy becomes critical.

Condition 3 tokens are identified as alveolar more frequently than condition 4 tokens, which, according to the EPG record, are completely assimilated. This evidence suggests that there is, in many if not all of the condition 3 tokens, some kind of articulatory 'residue' which is having acoustic consequences without leaving a trace in the EPG record. The question then arises: what is the nature of this acoustic cue, and how do phoneticians set about exploiting it in a phonetic transcription?

FIGURE 3

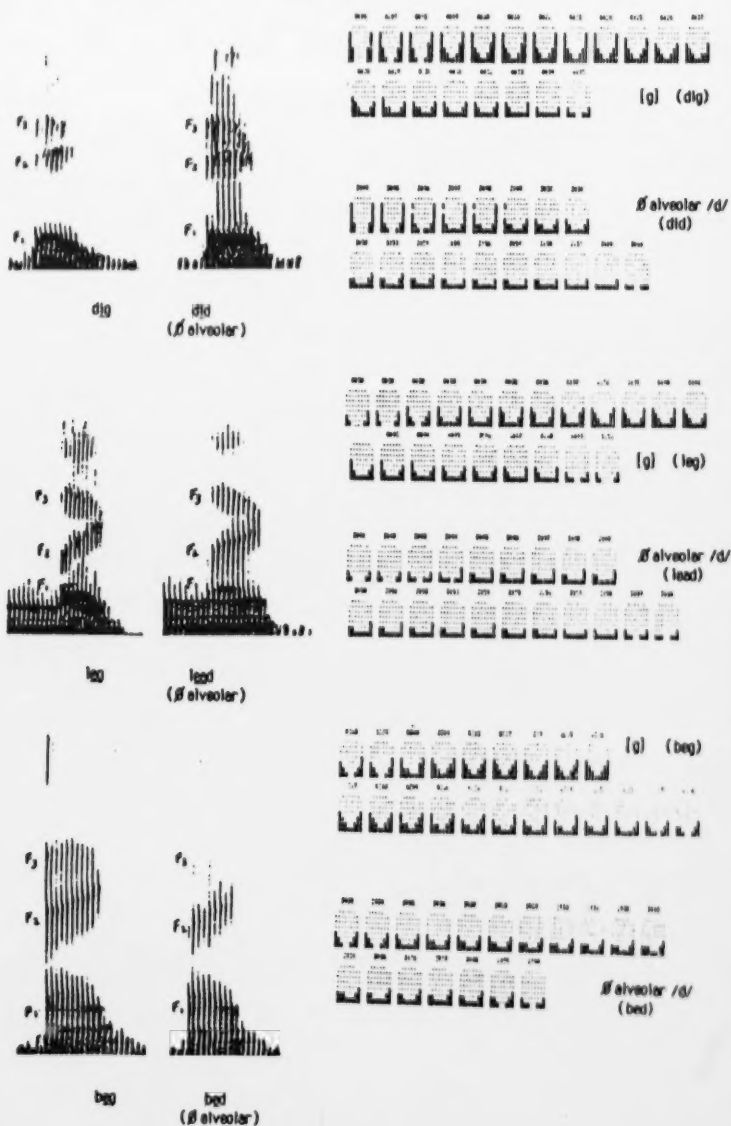
Percentage alveolar identifications for four EPG conditions



We will look first to see if EPG can give us any indications as to what these cues are. Remember that a residual alveolar contact shows up as lateral contact and partial alveolar contact. However, looking at some of the tokens that we originally classed as 'completely assimilated', we note something peculiar. This shown in Fig. 4 (overleaf). Note how, in these pairs, it is the assimilated alveolar that has the lesser lateral contact and the more retracted velar articulation. This is intuitively unexpected. But look at the identification scores for these three items (Table 4, overleaf).

For two of these three pairs (*lead/leg* and *bed/beg*), the difference between the scores for the two types is very much greater than for all the pairs taken together; as Table 4 shows, this is not true of any other single pair. There is, therefore, something differentiating these pairs rather

FIGURE 4
Palatograms and spectrograms showing alveolars and velars



clearly. This is evidently not alveolar contact. The EPG patterns of Fig. 4 can in fact be taken as evidence of a residual tongue body configuration appropriate for an alveolar: as the tongue tip moves up towards the alveolar ridge, the blade and pre-dorsum become concave; this reduces the amount of lateral contact in the pre-velar area. At the same time, this tongue shape will cause the velar contact itself to be more retracted. Some support for this interpretation is provided by spectrographic data

TABLE 4

Identification scores for individual tokens - EPG conditions 3 and 4

EPG CONDITION 3

Identifications as:

	<i>Alveolar</i>	<i>Non-alveolar</i>	<i>%alveolar</i>
did/dig	16	12	57
lead/leg	15	13	54
bed/beg	16	12	57
road/rogue	16	12	57
Byrd/Berg	12	16	43
fad/fag	8	20	29
bride/bribe	15	13	54

EPG CONDITION 4

Identifications as:

	<i>Alveolar</i>	<i>Non-alveolar</i>	<i>%alveolar</i>	Difference between conditions 3 & 4
did/dig	25	31	45	12
lead/leg	8	48	14	<u>40</u>
bed/beg	12	44	21	<u>36</u>
road/rogue	28	28	50	7
Byrd/Berg	27	29	48	-5
fad/fag	8	48	14	15
bride/bribe	24	32	43	11

for at least one of these pairs: in Fig. 4, the locus of F2 and F3 is higher for *dig* than for *did*, which suggests velar and alveolar offsets, respectively. This lingual configuration may in fact be heard as 'alveolarity'. This is the reason why the more retracted articulation is heard as more alveolar: it is the overall configuration of the tongue that has the acoustic consequences.

3.3 Transcription strategies - a mixed bag

EPG gives us, then, a clue as to the articulatory correlates of assimilation. In another paper (Wright & Kerswill, 1989), we have argued that this data suggests that there may be no such thing as 'complete' assimilation: there is always some articulatory 'residue' in 'maximally' assimilated items. However, here we shall look in some detail at how transcribers set about rationalising and reducing to symbols the differences they have heard. Table 5 shows the the transcriptions of condition 4 and condition 3 tokens of the three pairs just mentioned. (We have included only those transcriptions where (a) condition 4 tokens were correctly identified as velars, and (b) condition 3 tokens were correctly identified as alveolars and judged as having either partial or zero (but not full) alveolar articulation.) A striking overall pattern is the high frequency with which condition 3 is 'heard' as a partial alveolar rather than as the 'correct' zero alveolar. This should not surprise us, since once transcribers have decided they are listening to an alveolar, they will presumably try to indicate some sort of alveolarity in the transcription.

It is more interesting, however, to try to establish the strategies transcribers use to differentiate the velar and the alveolar tokens, and then to try to match these with the acoustic and articulatory data. An inspection of Table 5 shows there is much individual variation. However, three strategies seem to recur: these involve marking differences in vowel or consonant length, differences in vowel quality, and consonantal differences. We will discuss these in turn.

TABLE 5
 Transcriptions of tokens of three pairs of items

NOTE: only those transcriptions have been included where (a) condition 4 tokens were correctly identified as velars, and (b) condition 3 tokens were correctly identified as alveolars and judged as having either zero ('3') or partial ('2') (but not full) alveolar articulation.

Item:		Transcriber:
<u>1 did/dig</u>		
EPG	Judgm	A C E F
4 /g/	4	ɪg [~] ɪg [~] (3) ɪg (3) ɪg
3 /d/	3	ɪg [~] ɪg [~] (3) ɪg
3 /d/	2	ɪdg (2) ɪd (2) ɪdg
		H I J
4 /g/	4	ɪg ɪg (2) ɪg (4) ɪg ⁺ ɪg (3)
3 /d/	3	ɪg ɛg ɪg
3 /d/	2	ɪ(d)g ɪ(d)g ɪdg
<u>2 lead/leg</u>		
4 /g/	4	A B C D
4 /g/	4	ɛg [~] (3) ɛg (1) ɛg [~] ɛg [~] (2) ɛg [~] (2) ɛg [~] ɛg
3 /d/	3	ɛg [~] (3) ɛg [~] (1) ɛg [~] ɛg [~] (2) ɛg [~] (2) ɛg [~] ɛg
3 /d/	2	ɛdg ɛg, d ɛd [~] (2) ɛ(d)g ɛ(d)g
		F G H
4 /g/	4	ɛg (2) ɛg: ɛg: ɛg [~] (3) ɛg (3) ɛg
3 /d/	3	ɛg ed ɛ(d)k ɛ(d)g
3 /d/	2	ɛg ed ɛ(d)k ɛ(d)g

		I	J		
4 /g/	4	ɛg ɛg̃ (2)	ɛg ɛ̃g ɛ̃g ɛg		
3 /d/	3		ɛdg ɛ̃dg		
3 /d/	2	ɛg̃			
3 <i>bed/be</i>		A	C	D	F
4 /g/	4	ɛg̃ (4)	ɛg̃ ɛ̃g̃ (2)	ɛg (4)	ɛg̃ (2) ɛg
3 /d/	3	ɛ̃g̃	ɛ̃dg̃ (2)	ɛ̃g (2)	ɛ̃dg̃
3 /d/	2	ɛ̃dg̃			
4 /g/	4	H	I		
3 /d/	3	ɛ̃g̃ (4)	ɛ̃g̃ ɛ̃g̃ ɛ̃g̃ (2)		
3 /d/	2	ɛ̃g̃ ɛ̃g̃	ɛ̃g̃ ɛ̃g̃		

————— vowel quality difference
 - - - - - length difference

Vowel and consonant length

In five cases (enclosed in the table by a broken line), transcribers mark length differences. In three of these cases, the alveolar is heard as being preceded by a longer vowel, while in the other two the velar is given a longer consonant closure. Surprisingly, there is no evidence at all of longer vowels in *bed* and *did* than in their velar counterparts (see Fig. 4); yet for *lead*, whose vowel is measurably longer, no transcribers indicate this length. Consonant length differences (enclosed by a continuous line) can perhaps be seen as the other side of the same coin: a consonant after a durationally short vowel may be auditorily longer than after a longer vowel. If this is so, it is no less 'correct' to indicate a long consonant than to indicate a short vowel.

All five cases of length difference seem, then, to point in the same direction. However, there is disagreement between the transcribers as to

where this length resides. And where there is a clear vowel duration difference, this is apparently not heard as such; conversely, when there is no measurable difference in vowel duration, some transcribers seem to want to mark one. Whether or not there are consonant duration differences will have to await further spectrographic analysis. But for the moment, how can we explain the evident mismatch between measurable vowel durations and the transcriptions? As linguists, phoneticians 'know' about allophonic vowel duration differences, and it may be that they are trying to 'hear' such a difference - even though none is predicted phonologically (both /g/ and /d/ are voiced). Indicating length may be a more or less conscious attempt to rationalise a difference they can hear, using the limited resources of the IPA - one of which is to mark length. Alternatively, the percept of a length difference may be the psychoacoustic correlate of a consistent phonetic difference. As such, the percept is 'real' in the sense that it is not the consequence of an attempt to mark a difference willy-nilly, as in the case of the first explanation. Both explanations may have an element of truth in them: the fact that the differences marked by the transcribers are consistent with each other suggests a 'real' perceptual difference, while the disagreement as to where the difference lies suggests ad hoc attempts to indicate it using the transcription resources available.

Vowel quality differences

In twelve cases, we find vowel quality differences. In nine, the vowel before the velar is heard as closer than that before the alveolar; in only one case is the opposite true. Inspection of the spectrograms in Fig. 4 does not reveal any decisive differences; however, 'reading' vowel quality from the rapidly changing patterns on a spectrogram is notoriously difficult. There is obviously considerable agreement among the transcribers; even so, we must question the validity of their transcriptions because of the influence of their assumed prior 'knowledge' that closer allophones of vowels occur before velars. To test this source of error, we would need to carry out perception tasks using synthetic stimuli, or using edited natural stimuli from the which vowel offsets have been removed. However, the strength of the agreement certainly suggests the preservation of allophonic height differences even after the final consonant has

been apparently assimilated.

Consonant differences

In most of the cases, the transcribers note consonantal differences. This is particularly true, of course, where the transcriber has judged the alveolar as having a 'partial' articulation. There is a multiplicity of transcription strategies, suggesting that it is in the transcription of the consonant that the IPA itself fares worst. Strategies include using:

- 'no release' diacritic
- 'voicelessness' diacritic
- 'double articulation' diacritic
- 'retraction' diacritic
- 'fronting' diacritic
- 'length' diacritic
- 'shortness' diacritic
- 'lowering' diacritic
- parentheses
- superscripts

Some of these can be interpreted as representing the same intention on the part of the transcriber, though some can be taken simply to mean uncertainty (especially parentheses and superscripts). It is quite clear, however, that, unlike in the case of the vowels, the transcribers are explicitly aiming to represent *articulation* rather than, say, an abstract auditory parameter that might be labelled 'alveolarity'. The success of their enterprise will depend at the very least on (a) their ability to discriminate without being influenced by their phonological knowledge; (b) their experience with transcription; and (c) their knowledge of articulatory phonetics. To this must, of course, be added their degree of commitment to the task.

In representing what they hear for the consonants, the transcribers are constrained by the segmental nature of the IPA, and the relative difficulty of indicating phonetic features which change gradually over time and which are spread over more than one 'segment'.

4. Discussion

We think the identification of these tokens as alveolar or otherwise involves an extremely complex set of factors. Firstly, the listener must have knowledge of the auditory effects of different tongue gestures, including the 'residual' ones we are hypothesising. Secondly, as is well known, vowel quality and vowel length vary in different consonantal contexts; it is likely that these differences remain even after so-called assimilation has taken place, and continue to cue alveolarity. Lastly, an important part of these allophonic differences in vowels is that, in spite of certain universal tendencies, they are to a great extent dialect-specific, and the listener needs to have knowledge of the dialect (in this case, the speaker had mild south Yorkshire accent), and even knowledge of the speaker himself, to be able to unravel all these effects in such a way as to utilise them.

Transcription is a messy thing. For some people in this study, it is a way of representing a sequence of segments which are either articulatorily complete or non-existent - as some of the transcriptions show. Others seem more willing to allow incomplete or overlapping segments, but are still bound by the notion of articulatory segments. Yet others transcribe vowel quality differences. But we still don't know whether the vowel differences are due to residual articulatory gestures, or whether they are phonologically-determined, accent-specific allophonic differences that remain even where there is no residual articulatory gesture. In some cases, the transcribers could even, consciously or unconsciously, be tuning in to formant transitions which are not normally considered part of vowel quality and which are certainly not considered part of a phonological analysis.

To sum up, the problem lies in an inherent multi-layered ambiguity in the task of transcription itself. First, transcription is either meant to represent articulations, or it is meant to represent auditory impressions. Second, it either represents discrete segments, in which case it presupposes a prior phonological analysis, or it represents a continuously varying acoustic signal. Lastly, the continuous nature of the acoustic signal is either the result of pure, universal coarticulation or it is the result of

accent-specific allophonic and sandhi rules. The snag is, all these things are true to different degrees, and unfortunately transcribers will put the boundary between each of the pairs of opposites in different places. This is what we meant when we said at the outset that transcribing without any kind of theory is a dangerous thing: we simply do not know exactly what each individual is doing, and consequently we cannot interpret precisely what they write down.

REFERENCES

- Dressler, W U and Wodak, R. 1982. Sociophonological methods in the study of sociolinguistic variation in Viennese German. *Language in Society* 11.339-370.
- Harris, J. 1986. Phonetic constraints on sociolinguistic variation. *Sheffield Working Papers in Language and Linguistics* 3.120-43.
- Kerswill, P.E. 1984. Social and linguistic aspects of Durham *e*. *Journal of the International Phonetic Association* 14.1.13-34.
- Kerswill, P.E. 1985a. *A sociolinguistic study of rural immigrants in Bergen, Norway*. Unpublished Ph.D. dissertation. University of Cambridge.
- Kerswill, P.E. 1985b. A sociophonetic study of connected speech processes in Cambridge English: an outline and some results. *Cambridge Papers in Phonetics and Experimental Linguistics* 4.
- Kerswill, P.E. 1987. Levels of linguistic variation in Durham. *Journal of Linguistics* 23.25-49.
- Knowles, G.O. 1978. The nature of phonological variables in Scouse. In Trudgill, P. ed. 1978. *Sociolinguistic Patterns in British English*. London: Arnold.
- Lodge, K. 1986. *Studies in the Phonology of Colloquial Englishes*. London: Croom Helm.
- Milroy, J. 1976. Length and height variations in the vowels of Belfast vernacular. *Belfast Working Papers in Language and Linguistics* 13.
- Milroy, L. 1980. *Language and Social Networks*. Oxford: Blackwell.
- Milroy, L. 1987. *Observing and Analyzing Natural Language*. Oxford: Blackwell.
- Nolan, F.J. and Kerswill, P.E. 1989. The description of connected speech processes. In S. Ramsaran. ed. *Essays in Honour of A.C. Gimson*.

THE LIMITS OF AUDITORY TRANSCRIPTION

London: Routledge.

Orton, H., Halliday, W.J., Sanderson, S., Tilling, P.M., Wakelin, M.F. and Wright, N. 1962-7. *Survey of English Dialects* Vols. I-IV. Leeds: E.J. Arnold.

Orton, H., Sanderson, S. and Widdowson, J. 1978. *Linguistic Atlas of England*. London: Croom Helm.

Romaine, S. ed. 1982. *Sociolinguistic Variation in Speech Communities*. London: Arnold.

Trudgill, P. 1983. *On Dialect*. Oxford Blackwell.

Wright, S. 1986. The interaction of sociolinguistic and phonetically-conditioned CSPs in Cambridge English: auditory and electropalatographic evidence. *Cambridge Papers in Phonetics and Experimental Linguistics* 5.

Wright, S. and Kerswill, P. E. 1988a. *On the perception of connected speech processes*. Paper given at the Linguistics Association of Great Britain meeting. University of Durham. March 1988.

Wright, S. and Kerswill, P. E. 1988b. *EPG in the analysis of connected speech processes*. Paper given at the First National Symposium on the Clinical Applications of Electropalatography. University of Reading. March 1988.

Wright, S. and Kerswill, P. E. 1988c. *On the perception of connected speech processes*. Paper given at the Colloquium of the British Association of Academic Phoneticians, Trinity College. Dublin. March 1988.

Wright, S. and Kerswill, P.E. 1989. Electropalatography in the analysis of connected speech processes. *Clinical Linguistics and Phonetics* 3.

Wright, S. and Kerswill, P.E. in preparation. *On the perception of connected speech processes*.